基于关联数据的机构知识库实体学术关系发现方法研究[1]

梁美宏[1] 胡晓辉[1] 束彭飞[1] 刘飞[2]
1 河海大学图书馆 江苏南京 210098
2 南京工业大学浦江学院图书馆 江苏南京 211200

摘要：[目的/意义]利用关联数据唯一性、链接性等特点，对机构知识库实体进行学术关系发现与关联组织，实现机构知识库资源语义聚集。[方法/过程]分析实体学术关系类型及特征、关联数据要求，提出学术关系关联方法及流程，利用河海大学机构知识库数据开展实证研究。[结果/结论]构建了基于关联数据的实体学术关系发现方法，发现了实证数据间的学术关系，并指出其应用价值。通过该方法可以实现机构库资源语义关联，提供基于关联数据的知识检索服务，满足用户知识需求。

关键词：机构知识库；关联数据；学术关系；发现方法
分类号：G250

1 引言

机构知识库（Institutional Repositories，IR）是存储与管理、传播与分享学术机构知识资源的重要平台，对于促进科研成果共享交流、提升科学家及机构学术影响力、完善成果产出组织管理等方面均有重要作用。全球首个机构知识库是美国麻省理工学院于 2002 年构建的 DSpace@MIT，经过二十余年发展，截至 2023 年 2 月，登记于开放存取知识库目录 OpenDOAR 的机构知识库达到 6012 个，其中中国大陆地区共有 65 个 OpenDOAR. Browse by country [EB/OL]. [2023-02-28]. https://v2.sherpa.ac.uk/view/repository_by_ country/.。

机构知识库存储着本单位产生的论文、专利、研究报告等多种类型、不同格式的成果资源，这些资源之间蕴含着丰富的关联关系。通过发现资源间存在的关系并使之相关联，实现资源语义集成组织与一站式发现，是机构知识库建设与持续发展的重要内容。关联数据（Linked Data）提出于 2006 年，能够通过关联技术使分布在不同时间、不同空间的资源联系在一起，是机构知识库实现资源语义发现的重要技术手段。

学术关系是科学研究产生与发展过程中存在的关联关系，对机构知识库资源实体之间存在的学术关系进行发现关联，有利于揭示资源之间的学术联系，增强检索结果学术相关性，提高用户检索效率，促进机构知识库语义检索、智能导航、科学评价等深层次知识服务。本文试图在已有研究的基础上，探析基于关联数据的机构知识库实体学术关系发现方法，为机构知识库资源集成组织与知识发现提供参考。

2 相关研究现状

2.1 机构知识库实体学术关系研究现状

国内外已开展机构知识库实体关系方面的研究，包括①基于数据模型的关系关联研究，如孙翌等研究了机构库实体数据关联模型的构建孙翌,胡爱.基于多维度关联的机构知识库数据模型的构建与分析[J].现代情报,2018,38(07):95-106,159.，龚晓阳等分析了 CRIS 中 CERIF 模型的实体关系及其语义关联，可与机构知识库实现有机链接龚晓阳,刘啸,邵波.CRIS 技术发展及应用对中国高校机构知识库建设启示[J].图书馆学研究,2019(10):38-44,67.，Farid, H

---

1 本文系江苏省图书馆学会一般课题“机构知识库科研实体学术关系网络构建研究”（项目编号：22YB071）；江苏高校哲学社科一般项目“基于关联数据的高校机构知识库资源实体关系发现研究”研究成果之一。
作者简介：梁美宏（0000-0003-2321-9059），馆员，硕士，E-mail：liangmh188@126.com；胡晓辉，副研究馆员，硕士；束彭飞，馆员，硕士；刘飞，助理馆员，本科。

等研究了机构知识库实体与本体的语义关联 Farid H, Khan S, Javed M Y. DSont: DSpace to Ontology Transformation[J]. Journal of Information Science,2016(42):179-199.；②基于特定资源实体的关系关联研究，如陈嘉勇等研究了机构知识库文献数据实体关系陈嘉勇,周婕,李玲,等.基于文献实体关系模型的高校机构知识库作者认领模式研究[J].情报理论与实践,2015,38(02):59-63.，魏来等研究了机构知识库资源与书目数据的关系关联方法魏来,宁子晨.基于 RDA 的图书馆书目数据与机构知识库数据关联研究[J].图书馆工作与研究,2018(07):43-49.，Hallo, M 等研究了开放获取期刊关联数据发布方法，可应用于机构知识库数据关联组织 Hallo M, Lujan-mora S, Trujillo J. An Approach to Publish Statistics from Open-Access Journals Using Linked Data Technologies[C]. 9th International Technology, Education and Development Conference (INTED). INTED Proceedings. SPAIN: IATED-INT Assoc Technology Education & Development,2015:5940-5948.。

学术关系是指产生于学术生产与发展过程中的关联关系，有助于厘清学术发展的历史脉络、总结学术合作关系形成与发展的基本规律刘绍怀.学术链:客观存在的学术关系形态[J].思想战线,2011,37(01):1-3.。图书情报领域对学术关系的研究主要是基于社会网络分析、多元统计分析等方法，分析知识脉络发展、学术交流行为等学术关系问题任瑞娟,濮德敏,张媛.基于五维学术关系发现的知识脉络可视化实践[J].大学图书馆学报,2016,34(01):69-75.。专门研究机构知识库实体学术关系关联与发现的内容较少，本文作者及其团队在前述研究中研究并探讨了机构知识库科研实体学术关系发现体系孙清玉,梁美宏,胡晓辉.机构知识库科研实体学术关系发现体系研究[J].情报杂志,2022,41(11):179-184.。

2.2 关联数据在机构知识库中的应用现状

关联数据自产生以来，引起国内外广泛研究与应用。关联数据的本质是将各种分布式、异构数据通过语义关联连接在一起，核心理念是开放、关联、复用。2007 年万维网联盟 W3C 启动 LOD 关联开放数据（Linking Open Data）项目，并发布关联开放数据云图，截至 2023 年 2 月，其收集的开放关联数据集已达到 1594 个 The Linked Open Data Cloud. [EB/OL]. [2023-02-20].https://www.lod-cloud.net/.。图书馆界，瑞典国家图书馆最先将本国国家联合目录转换为关联数据，之后美国、英国、法国等国陆续将本国图书馆书目数据发布为关联数据张海玲. 图书馆书目数据的关联数据化研究——以德国国家图书馆为例[J]. 图书馆论坛,2013,33(1):120-125.。我国中国科学院图书馆、上海图书馆、中国科学技术信息研究所等多家机构及学者对关联数据保持较为持续的关注与研究，并形成应用成果，例如上海图书馆依托本馆资源以关联数据的方式发布了一系列知识库 尚渡新,袁润,夏翠娟,等.关联数据在知识库中应用的研究综述[J].数字图书馆论坛,2022(03):22-31.。

在机构知识库领域，国内学者的相关研究主要集中在以下几个方面。首先，关联数据在机构知识库构建中的应用，如周宇、欧石燕提出了一种基于关联数据的机构知识库构建方法周宇,欧石燕.面向关联数据的高校机构知识库构建方法研究[J].图书情报工作,2016,60(01):105-113.；郭卫兵、臧莉娟依据关联数据基本原则研究了国防科研机构知识库的构建与服务模式郭卫兵,臧莉娟.基于关联数据技术的机构知识库构建与服务[J].兵器装备工程学报,2020,41(12):275-280.。其次，关联数据在机构知识库数据关联与服务中的应用，如陈洁等探讨了科技报告数据与机构知识库的关联机制陈洁,韩非,武茜,等.科技报告数据关联机制研究[J].数字图书馆论坛,2017(01):46-50.；林静等研究了基于关联数据的机构知识库延伸服务林静,陈和,陈娟,等.基于关联数据消费的高校图书馆延伸服务探索——以厦门大学图书馆为例[J].大学图书馆学报,2020,38(03):71-79.。再次，关联数据在机构知识库资源发现中的应用，如王思丽、祝忠明研究了关联数据在机构知识库语义拓展中的应用，以促进机构库的资源发现服务王思丽,祝忠明.利用关联数据实现机构知识库的语义扩展研究[J].现代图书情报技术,2011(11):17-23.；都平平等在其研究中提到机构知识库的科学研究数据可通过关联数据建立关系链接都平平,李雨珂,孟勇,等.新一代机构知识库中研究数据的关联组织研究[J].现代情报,2018,38(12):86-90.；赵夷平研究了基于关联数据的机构知识库资源聚合与知识发现体系构建等方面内容赵夷平.基于关联数据的机构知识库资源聚合与知识发现研究[D].吉林大学,2018.。

综上所述，目前学术界已开展机构知识库资源实体关系、关联数据在机构知识库中的应用等相关研究，但深入分析实体关系，探析基于关联数据技术的机构知识库实体学术关系关联与发现的研究尚未有之。本文研究的主要目标是：其一，基于机构知识库科研实体之间存在的学术关系，提出利用关联数据技术的机构知识库实体学术关系发现方法；其二以河海大学机构知识库为数据源，按照本文提出的方法进行实证研究，验证方法的可行性有效性。

3 基于关联数据的学术关系发现方法

利用关联数据进行学术关系发现，首先要明晰机构知识库资源实体之间存在的学术关系类型有哪些，以及学术关系涉及到的特征属性；其次要将资源数据转换为符合关联数据格式的数据；再次基于合适的关联方法，发现数据学术关系，并建立关联链接；最后通过可视化方法，揭示学术关系关联结果，实现机构知识库资源的语义聚集。

3.1 学术关系类型

本文作者及其团队在先前研究中，通过分析机构知识库主要资源类型及其特征属性，以学者、机构、学科、资源作为分析对象，依据学术关系内涵，将机构知识库主要科研实体的学术关系分为师承关系、合作关系、引证关系、贡献关系、附属关系等类型孙清玉,梁美宏,胡晓辉.机构知识库科研实体学术关系发现体系研究[J].情报杂志,2022,41(11):179-184.。

具体来看，学者层面，学者与学者之间存在师承、合作、引证等学术关系；学者与机构之间存在隶属关系；学者与学科之间存在研究领域隶属关系；学者与资源之间存在贡献关系。机构层面，机构与机构之间存在合作关系，包含一级机构、二级机构不同层级机构间的合作关系；机构与学者之间存在隶属关系；机构与学科之间存在研究领域隶属关系；机构与资源之间存在贡献关系。学科层面，学科与学科之间存在前沿热点、学科交叉等学术关系；学科分别与学者、机构、资源之间存在研究领域隶属关系。资源层面，资源之间存在引证、附属、相关等学术关系；资源与学者之间存在被贡献关系；资源与机构之间存在隶属关系；资源与学科之间存在研究领域隶属关系孙清玉,梁美宏,胡晓辉.机构知识库科研实体学术关系发现体系研究[J].情报杂志,2022,41(11):179-184.。各实体间的学术关系及其对应资源属性情况如表 1 所示。

表 1 实体主要学术关系类型

实体	学术关系	对应资源	涉及特征属性	意义
学者	师承关系	学位论文为主	作者；导师	构建学术谱系，梳理学术脉络
	合作关系	含两个或两个以上作者的资源	作者	了解学者间学术交流、学科交叉、团队构成等情况
	引证关系	含参考文献的资源	作者；参考文献作者	通过共被引、耦合、互引等关系分析，揭示学者间知识交流分布规律
机构	合作关系	含两个或两个以上机构的资源	机构	了解机构间学术交流合作情况
学科	前沿热点	文本资源为主	关键词；主题词；分类号	通过共现分析、聚类分析等，了解学科研究前沿、热点
	学科交叉	文本资源为主	关键词；主题词；分类号	通过共现分析等，发现学科交叉点

资源	引证关系	含参考文献的资源	资源标识符；参考文献标识符	通过引证分析，探索学术知识流动、学术演进态势
	附属关系	文本资源；数据资源；实体资源等	来源；资源类型；项目名称等	掌握完整的学术研究资料、数据、成果
	相关关系	所有类型资源	作者；机构；主题词；分类号等	按相关程度组织具有一定相关关系的学术研究成果

3.2 关联数据准备

利于关联数据技术对机构知识库科研实体关系进行关联与发现，前提是将机构知识库资源数据转换为符合关联数据格式的数据。依据关联数据概念及四个基本原则，可以将关联数据准备过程分为以下三个步骤。

第一，数据描述与标准化处理。机构知识库存储的资源内容既有人工上传的数据、也有机器抓取数据，其数据结构内容各不相同，需要通过标准化描述实现格式统一，包括构建元数据集、构建规范文档、属性描述方案等。其中所构建元数据集主要是机构知识库存储的不同类型资源元数据集；所构建规范文档应至少包括：学者规范文档、机构规范文档学科规范文档；数据属性描述方案可选择都柏林核心元数据（Dublin Coremetadata, DC 元数据）方案。需要注意的是，机构知识库在建设期间通常已建有部分基于特定元数据方案的元数据集、学者词典、机构词典等，因此在进行数据标准化处理时，要考虑已有内容的复用或映射。

第二，数据 RDF 转化与 URI 命名。关联数据是以 URI 命名、RDF 三元组格式存储的数据梁美宏.书目关联数据的关联方法研究[D]. 北京:中国科学技术信息研究所,2016.，因此要将标准化后的机构知识库数据转换为 RDF 形式、URI 命名方式。其中 RDF 不仅可以提供数据间的联系，而且携带语义信息，便于数据识别和使用，通常一个 RDF 三元组由资源、属性、属性值组成，表示资源的一个属性 W3SCHOOL. RDF[EB/OL]. (2012-09-28) [2023-02-18]. http://www.w3 school.com.cn/rdf/rdf_intro.asp.；统一资源标识符（Uniform Resource Identifier，URI）是互联网资源的标识符体系,用来对能够被识别的任何电子资源的名称或地址进行唯一标识 BERNERS-LEE T,FIELDING R,MASINTERL. Uniform resource identifiers (URI):generic syntax[EB/OL]. (1998-08-01) [2023-02-18]. http://www. ietf.org/rfc/rfc2396.txt.，URI 命名时既可复用已有资源、也可依据相关语法规则自定义标识符。经过转化与命名的机构知识库资源与实体得到了唯一化标识，并可建立链接，例如资源作者属性可表示为 scholar:{<URI>,dc:creator,name}，学者所属机构属性可表示为 institution:{<URI>,dc:institution,name}。机构知识库实现关联数据转化的内容至少包括资源、学者、机构、学科，即这几个实体应拥有唯一标识符。

第三，关联数据发布。目前关联数据发布平台有多种，机构知识库资源在进行关联数据发布时，应根据资源规模、数据更新频率等情况，采用适合本机构库实情的方式，选择适当的平台将规范化的机构知识库数据发布为关联数据，为知识发现服务提供资源保障。

3.3 学术关系关联方法

基于上文分析的机构知识库主要科研实体间学术关系类型及相关资源特征属性，采用属性值匹配的推理方法，以发现机构知识库资源、学者、机构、学科实体之间的学术关系，在选择属性时，优先选择唯一标识符，在没有唯一标识符的情况下，选用相对规范的属性，如名称、编号、主题词、分类号等。具体关联方法及过程如图 1 所示。

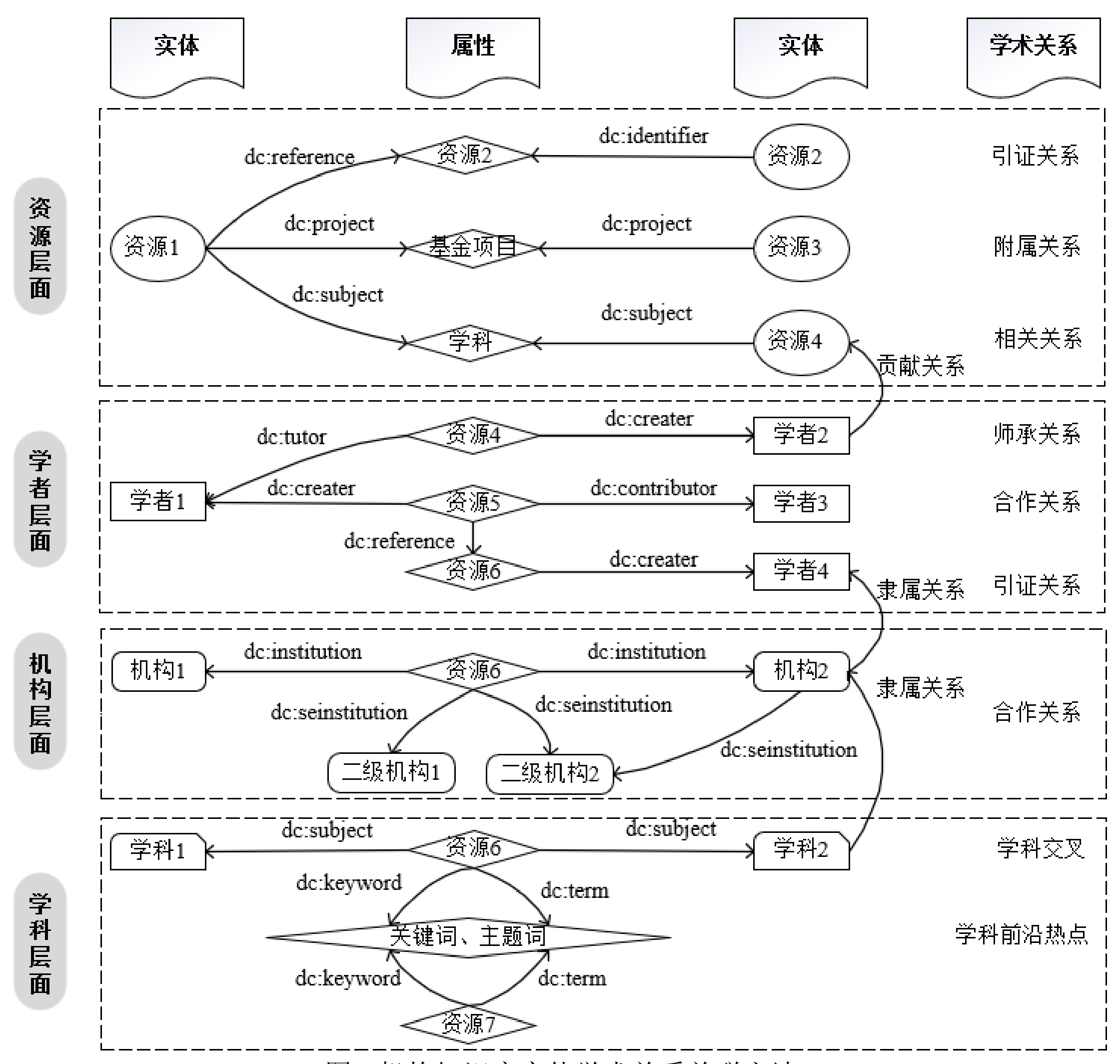


图 1 机构知识库实体学术关系关联方法

资源层面，依据参考文献、唯一标识符、所属项目、学科等属性，判断资源之间学术关系。通过匹配 URI 唯一标识符属性值是否相同，确认引证关系，主要是资源及参考文献的 URI；匹配资源来源、类型、所属项目等属性值是否相同，确认附属关系，图中以所属基金项目为例进行说明；匹配作者、机构、主题词、学科等属性值是否相同，确认相关关系，图中以学科为例进行说明。

学者层面，依据作者、导师、其他作者等属性，判断学者之间学术关系。通过匹配同一资源作者和导师属性值，确认师承关系；匹配同一资源作者和其他作者属性值，确认合作关系；匹配资源作者和参考文献作者属性值，确认引证关系。

机构层面，依据机构属性，判断机构之间学术关系。通过匹配同一资源机构属性值，确认一级机构合作关系。在相同一级机构下，匹配二级机构属性值，确认二级机构合作关系。

学科层面，依据学科、关键词、主题词等属性，判断学科之间学术关系。通过匹配资源学科属性值，确认学科交叉关系；匹配资源关键词、主题词等属性，通过社交网络分析等途径得到学科前沿热点。

3.4 学术关系发现结果

根据学术关系关联方法及其流程，可以得到机构知识库资源、学者、机构、学科实体之间的学术关系，通过数据 URI 唯一标识符使资源实现语义链接，利用合适的可视化工具使具有学术关系的可链接数据得到可视化展示，有利于资源语义聚合组织，便于用户浏览与获取信息内容。经过发布后的关联数据能够使机构知识库资源与分布在网上的其他资源进行关联，扩大学术关系网络语义组织，为学术交流与扩散提供帮助。

4 基于关联数据的学术关系发现实证

河海大学机构知识库建设以来，较全面地收集了本校在职教师的中外文科研成果，并建有元数据框架、作者表和机构表。本文以该机构知识库为数据来源，依据上文研究的学术关系发现方法开展实证研究，验证所构建方法的科学性、合理性、有效性。

4.1 数据获取与处理

在机构知识库中选择 10 条数据作为实证样本数据，开展实证研究。数据处理流程包括：导出数据；构建元数据集，并进行初步清洗加工；在机构库已有的元数据框架、作者表、机构表等基础上，构建学者规范文档、机构规范文档、学科规范文档；采用 DC 元数据方案对数据进行规范描述；将规范描述好的数据转化为 RDF 格式，并以统一资源标识符 URI 命名。最终得到关联数据格式的数据。

（1）规范文档建立与 DC 描述

第一构建元数据集。将 10 条资源数据导出为结构化的二维表格式，使表中每一列表示资源的一个属性，每一行为一个资源的所有属性值。在 Excel 中初步清洗数据，包括将其他作者属性中的多名作者分列，使每个属性仅对应一个属性值；将机构属性中一级机构和二级机构名称分列存储，并去掉其中地址、邮编等非机构名称内容，其中二级机构仅存储河海大学的二级机构；统一中外文属性中的标点符号格式；补充重要属性的缺失数据等等最终构建出资源元数据集。

第二构建规范文档。学者规范文档以机构库建有的作者表和校人事处提供的教师信息表为基础进行构建，以河海大学师生为建设主体，其中缺失内容通过官方网站调研、学术文献中作者简介等途径予以补充。机构规范文档以机构库建有的机构表为基础进行构建，二级机构以河海大学二级机构为建设主体，缺失内容通过官方网站调研予以补充。学科规范文档以机构库建有的学科表为基础进行构建，并将教育部一级学科作为规范名称，中图分类号、ESI 学科、WOS 学科分别与教育部学科进行映射，并作为别称，其中我国目前最新的教育部学科分类共设置 117 个一级学科国务院学位委员会,教育部.研究生教育学科专业目录（2022 年）[EB/OL]. (2022-09-13) [2023-03-15]. http://www.gov.cn/zhengce/zhengceku/2022-09/14/5709785/files/bb23485679bf489aba5d3299c36e0606.pdf.，第五版中图分类法共分五大部类、22 个基本大类国家图书馆《中国图书馆分类法》编辑委员会.中国图书馆分类法(第五版)[M].北京: 国家图书馆出版社,2010.，Web of Science 数据库将 ESI 学科分为 22 大类、WOS 学科分为 252 个方向 Clarivate. Web of Science Research Areas [EB/OL]. (2021-12-01) [2023-03-15]. https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm?Highlight=research%20area.。表 2 列出了所构建规范文档部分属性和机构库已有词表字段的映射情况。基于所构建规范文档，将元数据集中的数据进行规范化清洗。

表 2 规范文档属性与机构库词表字段映射情况

学者规范文档属性	机构库作者表字段	机构规范文档属性	机构库机构表字段	学科规范文档属性	机构库学科表字段
唯一标识符	编号	唯一标识符	编号	唯一标识符	/
规范姓名	姓名	规范名称	机构名	规范名称	教育部一级学科
出生年月	/	成立时间	/	分类号	中图分类号
别名	曾用名	别称	曾用名	ESI 学科	ESI 学科
籍贯	/	所在地区	所属地区	WOS 学科	WOS 学科
性别	/	上级机构	上级机构		
所属机构	二级机构	下级机构	下级机构		
所属学科	/				
研究方向	/				

职称　　　职称		

第三 DC 元数据描述。DC 元数据方案包含 15 个核心元素，根据机构库已有元数据框架和学术关系匹配属性需求，扩展元数据元素，最终得到元数据方案如表 3。基于该方案对实证样本数据、所建规范文档数据进行 DC 元数据描述，使不同结构格式的数据得到统一描述。

表 3 机构库元数据描述方案

数据属性	**机构库元数据字段**	**DC 元数据描述**
唯一标识符	/	dc：identifier
标题	title_c	dc：title
其他题名	title_e	dc：otitle
第一作者	showwriter	dc：creator
其他作者	/	dc：contributor
导师	bstutorsname	dc：tutor
机构	showorgan	dc：institution
二级机构	/	dc：seinstitution
学科	/	dc：subject
关键词	keyword	dc：keyword
主题词	/	dc：term
参考文献	/	dc：reference
出版物	publisher	dc：publisher
出版时间	years	dc：date
资源类型	/	dc：type
语种	language	dc：language
卷期	vol；num	dc：volume
基金项目	fund	dc：fund

（2）RDF 转化与 URI 命名

将上述 DC 描述化后的数据转化为 RDF 格式，表示实体与属性的关系。依据 RDF 组成内容，二维表中每一行数据整体（一个实体数据）为一个 RDF 三元组主体，主键数据为主体的 URI 标识符，每一列标题（属性）为三元组谓词，每一列数据（属性值）为三元组客体。二维表行数表示实体个数，列数表示每个实体的 RDF 三元组数量。

数据 URI 命名包括对资源、学者、机构、学科的命名。以机构知识库已有数据 URI 为基础，命名机制为：前缀+唯一标识编号，其中资源、学者、机构、学科的 URI 命名前缀分别为："http://ir.hhu.edu.cn/Articles/Article_Detail.aspx?id="、"http://ir.hhu.edu.cn/writer/rw_zp.aspx?id="、"http://ir.hhu.edu.cn/organ/jg_zp.aspx?id="、"http://ir.hhu.edu.cn/class/db_zp.aspx?id="。

最终得到经过统一规范清洗与描述后的部分规范文档、实证样本数据分别如图 2、图 3 所示，图中唯一标识符均省略前缀内容。

序号	唯一标识符	规范姓名	出生年份	籍贯	性别	所属机构	所属机构	所属学科	研究方向	职称	别名	别名	别名	……
	dc:identifier	dc:name	dc:date	dc:birthplace	dc:gender	dc:seinstitution	dc:seinstitution	dc:subject	dc:research	dc:rank	dc:oname	dc:oname	dc:oname	……
1	1751e53621f78b45	李晓蕾	1998		女	水文水资源学院			水文循环响应研究	/	lixiaolei	li,xiaolei	li xiao-lei	……
2	8437a0c31de89168	王卫光	1979	河南开封	男	水文水资源学院	水文水资源与水利工程科学		生态水文;水文预报;水循环	教授	wangweiguang	wang,weiguang	wang wei-guang	……
3	80d89fd4e41759a1	薛联青	1973		女	水文水资源学院			环境水文;生态水文	教授	xuelianqing	xue,lianqing	xue lian-qing	……
5	1dff317764fc553e	朱世峰	1999		男	水文水资源学院			水文物理规律;水文预报	/	zhushifeng	zhu,shifeng	zhu shi-feng	……
6	21e12e6f9c57b8a3	周忠丽	1981	湖南岳阳	女	马克思主义学院			基层治理;基层党建	教授	zhouzhongli	zhou,zhongli	zhou zhong-li	……
7	235f663fdc021524	鲍建腾	1989		男	水文水资源学院			水利调度管理	/	baojianteng	bao,jianteng	bao jian-teng	……
8	246c8f8870e29ac8	宋瑞明	1989	河南濮阳	女	水文水资源学院			水文物理规律;水文预报	/	songruiming	song,ruiming	song rui-ming	……
9	27b766d08f77776b	张翔宇	1991	河南焦作	男	水文水资源学院			水资源规划与管理	/	zhangxiangyu	zhang,xiangyu	zhang xiang-yu	……
10	27c7d80c4c93f80c	郑巍斐	1992	浙江金华	男	水文水资源学院			水文水资源	/	zhengweifei	zheng,weifei	zheng wei-fei	……
11	585ca080934d963a	杨肖丽	1976	河北邯郸	女	水文水资源学院			水文循环	副教授	yangxiaoli	yang,xiaoli	yang xiao-li	……
12	7e145c1853f0fc2d	张增信	1977	山东青岛	男	水文水资源学院			气候变化;生态水文	教授	zhangzengxin	zhang,zengxin	zhang zeng-xin	……
13	2ad0a1122be07b04	陈泽峰	1994	江苏无锡	男	水文水资源学院	水文水资源与水利工程科学		生态水文;环境水文	/	chenzefeng	chen,zefeng	chen ze-feng	……
14	2c14ba5a80779224	傅健宇	1993	湖南怀化	男	水文水资源学院			生态水文;环境水文	/	fujianyu	fu,jianyu	fu jian-yu	……
15	4dd675f6625f3c88	邓超	1989	湖南常德	男	水文水资源学院			流域水文	讲师	dengchao	deng,chao	deng chao	……

（学者规范文档）

序号	唯一标识符	规范名称	成立时间	所在地区	上级机构	下级机构	下级机构	别称	别称	别称	别称	……
	dc:identifier	dc:name	dc:date	dc:area	dc:superior	dc:subordinate	dc:subordinate	dc:oname	dc:oname	dc:oname	dc:oname	……
1	21f30d5836da52f9	水文水资源学院	1952	江苏南京	河海大学	水资源研究所	地理信息科学与工程研	水文系	水资源环	College of	Coll Hydro & Water Resources	
2	b2f4bcfcde400cca	水利水电学院	1952	江苏南京	河海大学	水工结构研究所	智慧水利与智能减灾研	水工结构系	水利水电	College of	Coll Water & Hydropower Engn	
3	c1d3faa61e2d2d9c	港口海岸与近海工程学院	1952	江苏南京	河海大学	海岸带资源与环境	海洋可再生能源工程研	港口工程系	交通与海	College Ha	Coll habor Coastal & Offshore engn	
4	3a68fc59894b6689	土木与交通学院	1995	江苏南京	河海大学	结构工程研究所	安全与防灾工程研究所	土木工程学	College of	Coll Civil & Tranportat Engn		
5	1c97a5f0f0d172c2	马克思主义学院	2011	江苏南京	河海大学	马克思主义理论与	思想政治教育研究所	College of M	Coll Marxi	School of	Sch Marxism	
6	98466636dbe53b38	水文水资源与水利工程科学国家重点实验室	2004	江苏南京	河海大学			State Key La	State Key	Key Lab H	State Key Lab Hydrol & Hydraul Engn	

（机构规范文档）

序号	唯一标识符	规范名称	分类号	ESI学科	WOS学科	WOS学科	WOS学科	WOS学科
	dc:subject	dc:name	dc:class	dc:esi	dc:wos	dc:wos	dc:wos	dc:wos
1	CSC0101	哲学	B0;B1;B2;	social sciences, general	philosophy	ethics	logic	
2	CSC0201	理论经济学	F0	economics & business	economics			
3	CSC0202	应用经济学	F0	economics & business	business, finance			
4	CSC0301	法学	D9;DF	social sciences, general	law			
5	CSC0302	政治学	D0;D5;D6	social sciences, general	asian studies	international relations	area studies	
6	CSC0303	社会学	C0;C1;C2;	social sciences, general	sociology	social sciences, interdisc	anthropology	demography
7	CSC0304	民族学	C95	social sciences, general	ethnic studies			
8	CSC0305	马克思主义理论	A1;A5;A7;	social sciences, general	history			
9	CSC0306	公安学	D035.3	social sciences, general				
10	CSC0307	中共党史党建学	D2	social sciences, general				
11	CSC0308	纪检监察学	D2	social sciences, general				
12	CSC0401	教育学	G4	social sciences, general	education & education	education, scientific disc	education, special	
13	CSC0402	心理学	B84	psychiatry/psychology	psychology	psychology, applied	psychology, develop	psychology, mathem
14	CSC0403	体育学	G8	social sciences, general				
15	CSC0501	中国语言文学	I2	social sciences, general	cultural studies	language & linguistics	linguistics	literature
16	CSC0502	外国语言文学	I1;I3	social sciences, general	literature, african, aust	literary theory & criticis	literature, american	literature, romance
17	CSC0503	新闻传播学	G2	social sciences, general	film, radio, television			
18	CSC0601	考古学	K85;K86;I	social sciences, general				
19	CSC0602	中国史	K2	social sciences, general				
20	CSC0603	世界史	K1	social sciences, general				
21	CSC0701	数学	O1	mathematics	mathematics, applied	statistics & probability	mathematical & con	mathematics, interdis
22	CSC0702	物理学	O4	physics	physics, atomic, mole	physics, mathematical	physics, particles &	physics, condensed

（学科规范文档）

图 2 规范描述化的部分规范文档数据

序号	唯一标识符	标题	其他题名	第一作者	其他作者	其他作者	导师	机构	二级机构	二级机构	机构
	dc:identifier	dc:title	dc:otitle	dc:creator	dc:contributor	dc:contribu	dc:tutor	dc:instituti	dc:seinstitution	dc:seinstitution	dc:institution
1	675260aa3f254	基于CMIP6多模式的		1751e53621f78b45	8437a0c31de89168	张淑林		河海大学	21f30d5836da52f9	98466636dbe53b38	
2	ba3d2a034173	基于CMII	Attribution	1751e53621f78b45			8437a0c3	河海大学	21f30d5836da52f9		
3	b2a1aa31800e	基于CMIP5的塔里木		刘锋	80d89fd4e41759a1	魏光辉		河海大学	21f30d5836da52f9		新疆农业大学
4	8f1e08b99fa74	基于多种指标的长江		1dff317764fc553e	8437a0c31de89168	魏佳		河海大学	21f30d5836da52f9	98466636dbe53b38	
5	e34a3eef014f4	水工程精准调度的哲		21e12e6f9c57b8a3	235f663fdc021524			河海大学	1c97a5f0f0d172c2	21f30d5836da52f9	
6	30a3d174370e	江苏省水稻高温热害		246c8f8870e29ac8	8437a0c31de89168	27b766d08f77776b		河海大学	21f30d5836da52f9		
7	f2dac9e807494	基于统计降尺度的长		27c7d80c4c93f80c	程雪蓉	585ca080934d963a		河海大学	21f30d5836da52f9	98466636dbe53b38	江西省上饶市水文局
8	d5eada6fbf794	Assessment of CMIP		Tao, Hui	Gemmer, Marco	7e145c1853f0fc2d		河海大学	98466636dbe53b38		Chinese Acad Sci
9	9fad57eea05c4	中国区域植被对降水		2ad0a1122be07b04	8437a0c31de89168	2c14ba5a80779224		河海大学	98466636dbe53b38		
10	66daa1375bc64	Runoff Predicting and		4dd675f6625f3c88	8437a0c31de89168			河海大学	98466636dbe53b38	21f30d5836da52f9	

（1）

学科	学科	分类号	分类号	WOS学科	关键词	主题词	出版物	出版时间	资源类型	语种	卷期	基金项目	参考文献	参考文献	……
dc:subject	dc:subject	dc:class	dc:class	dc:wos	dc:keywor	dc:term	dc:publishe	dc:date	dc:type	dc:languag	dc:volume	dc:fund	dc:referenc	dc:referenc	……
水利工程	大气科学	TV125	P467		长江流域;CMIP6;气		中国农村	2022	期刊论文	英文		国家自然科学基金项目(51779073)	f2dac9e80	d5eada6fb	……
水利工程		TV12			蒸散发；气候变化情景；CMIP			2022	学位论文	中文					……
大气科学	水利工程	P426.61	TV125		气候变化;CMIP5;降		水电能源	2022	期刊论文	中文	40(3)	国家自然科学基金项目(U1603343)			……
农业资源与环境		S166			CN05.1;水稻;高温热		节水灌溉	2022	期刊论文	中文		国家自然科学基金项目(51979071)	30a3d174	675260aa3	……
水利工程		TV213.4			精准调度;水工程;哲		中国水利	2022	期刊论文	中文					……
农业资源与环境		S166			统计降尺度;气候变		灌溉排水	2017	期刊论文	中文	36(1)	国家自然科学基金面上项目(51379057)			……
大气科学	水利工程	P467	TV125		CMIP5;EDCDFm;降		中国农村	2017	期刊论文	中文	9	国家自然科学基金面上项目(51579066)			……
环境科学与工程				Environmental Sciences	Catchments;Watershe		CLIMATI	2012	期刊论文	英文	111(3-4)	国家自然科学基金项目(40771012)			……
大气科学	环境科学与工程	P467	X37		NDVI;植被生长;降水		灌溉排水	2019	期刊论文	中文	38(5)	国家自然科学基金项目(51779073)			……
环境科学与工程				Environmental Sciences	climate change; runof		SUSTAIN	2019	期刊论文	英文	11(21)	国家自然科学基金项目(51779073)			

（2）

图 3 规范描述化的实证样本数据

4.2 数据关联过程及关联结果

根据上文构建的学术关系关联方法及其过程，对实证样本数据进行属性值匹配推理，得到实体之间的学术关系，并建立链接。

为方便描述实体之间学术关系，图 4 以上述规范描述数据中的序号代表实体，以实体之间存在学术关系类型较多的资源序号为 1、2、8、10 四条数据为例，展示了数据大致匹配过程，例如资源 1 的参考文献为资源 8，因此二者存在引证关系；资源 1 与资源 2 的学科

属性相同，因此二者存在相关关系；等等。表 4 列出了实证样本数据中各实体之间的部分学术关系。

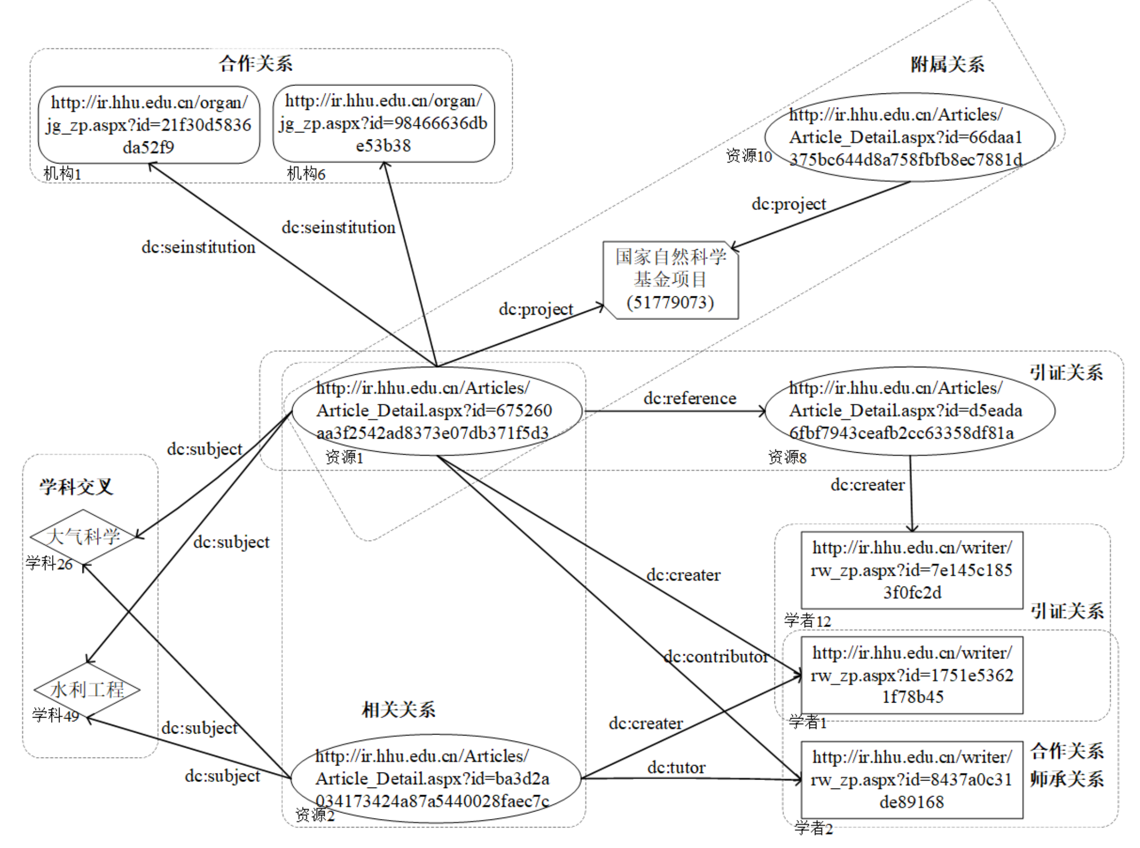


图 4 实体学术关系匹配情况

表 4 实证数据实体部分学术关系

资源序号	资源序号	学术关系	学者序号	学者序号	学术关系
1	7	引证；相关	1	2	合作；师承；引证
1	8	引证	1	10	引证
1	9	附属；相关	1	11	引证
1	10	附属	1	12	引证
1	2	相关	2	10	引证
1	3	相关	2	11	引证
1	5	相关	2	12	引证
2	3	相关	5	2	合作；引证
2	5	相关	5	8	引证
2	7	相关	5	9	引证
3	5	相关	2	8	引证；合作
3	7	相关	2	9	引证；合作
3	9	相关	5	1	引证
4	6	引证；相关	6	7	合作
4	1	引证	8	9	合作
5	7	相关	10	11	合作
7	9	相关	13	2	合作
8	9	相关	13	14	合作
8	10	相关	14	2	合作
9	10	相关	15	2	合作
机构序号	**机构序号**	**学术关系**	**学科序号**	**学科序号**	**学术关系**

1	5	合作	49	26	交叉
1	6	合作	26	64	交叉

在学术关系关联结果基础上，通过资源 URI 相互链接指向，能够使拥有学术关系的实体实现语义关联聚集，提高学术关系发现的深度与广度，利用可视化方法将实体及其学术关系展现出来，能够使用户更加便捷、高效地获取利用机构知识库存储的资源。

5 结语与展望

依据基于关联数据的学术关系发现方法，能够发现机构知识库不同实体之间存在的学术关系，例如资源之间引证关系、学者之间师承关系、机构之间合作关系等等，使学术生产过程中产生的实体与成果得以语义联结在一起，利于知识脉络梳理、学术发展创新。该方法应用于机构知识库资源组织建设，有助于机构知识库优化资源组织、扩展知识服务，促进机构成果理解与传播、交流与创新。

本文研究仍存在一些不足之处：例如学术关系类型的分类未必全面准确，需要在实践工作中结合数据内容不断完善；学术关系关联方法的科学性、合理性需要通过大量不同类型资源实证，逐步修正；学术关系发现结果的有效性、正确性需要进一步核验。今后将在研究中通过更多数据实证工作，结合理论与实践，进一步完善学术关系发现方法，使得所构建方法与发现结果更加具有应用价值。

参考文献

[1] OpenDOAR. Browse by country [EB/OL]. [2023-02-28]. https://v2.sherpa.ac.uk/view/repository_by_ country/.

[2] 孙翌,胡爱.基于多维度关联的机构知识库数据模型的构建与分析[J].现代情报,2018,38(07):95-106,159.

[3] 龚晓阳,刘啸,邵波.CRIS 技术发展及应用对中国高校机构知识库建设启示[J].图书馆学研究,2019(10):38-44,67.

[4] Farid H, Khan S, Javed M Y. DSont: DSpace to Ontology Transformation[J]. Journal of Information Science,2016(42):179-199.

[5] 陈嘉勇,周婕,李玲,等.基于文献实体关系模型的高校机构知识库作者认领模式研究[J].情报理论与实践,2015,38(02):59-63.

[6] 魏来,宁子晨.基于 RDA 的图书馆书目数据与机构知识库数据关联研究[J].图书馆工作与研究,2018(07):43-49.

[7] Hallo M, Lujan-mora S, Trujillo J. An Approach to Publish Statistics from Open-Access Journals Using Linked Data Technologies[C]. 9th International Technology, Education and Development Conference (INTED). INTED Proceedings. SPAIN: IATED-INT Assoc Technology Education & Development,2015:5940-5948.

[8] 刘绍怀.学术链:客观存在的学术关系形态[J].思想战线,2011,37(01):1-3.

[9] 任瑞娟,濮德敏,张媛.基于五维学术关系发现的知识脉络可视化实践[J].大学图书馆学报,2016,34(01):69-75.

[10] 冯靖雯,赵勇.学术谱系视角下杰出科学家的师承关系特征研究——以诺贝尔化学奖得主 Lipscomb 为例[J].情报工程,2020,6(06):22-32.

[11] 严炜炜,温馨.关注与合作：科研用户在线学术关系网络与行为模式分析[J].情报理论与实践,2022,45(04):75-82.

[12] 孙清玉,梁美宏,胡晓辉.机构知识库科研实体学术关系发现体系研究[J].情报杂志,2022,41(11):179-184.

[13] The Linked Open Data Cloud. [EB/OL]. [2023-02-20].https://www.lod-cloud.net/.

[14] 张海玲. 图书馆书目数据的关联数据化研究——以德国国家图书馆为例[J]. 图书馆论坛,2013,33(1):120-125.

[15] 尚渡新,袁润,夏翠娟,等.关联数据在知识库中应用的研究综述[J].数字图书馆论坛,2022(03):22-31.

[16] 周宇,欧石燕.面向关联数据的高校机构知识库构建方法研究[J].图书情报工作,2016,60(01):105-113.
[17] 郭卫兵,臧莉娟.基于关联数据技术的机构知识库构建与服务[J].兵器装备工程学报,2020,41(12):275-280.
[18] 陈洁,韩非,武茜,等.科技报告数据关联机制研究[J].数字图书馆论坛,2017(01):46-50.
[19] 林静,陈和,陈娟,等.基于关联数据消费的高校图书馆延伸服务探索——以厦门大学图书馆为例[J].大学图书馆学报,2020,38(03):71-79.
[20] 王思丽,祝忠明.利用关联数据实现机构知识库的语义扩展研究[J].现代图书情报技术,2011(11):17-23.
[21] 都平平,李雨珂,孟勇,等.新一代机构知识库中研究数据的关联组织研究[J].现代情报,2018,38(12):86-90.
[22] 赵夷平.基于关联数据的机构知识库资源聚合与知识发现研究[D].吉林大学,2018.
[23] 梁美宏.书目关联数据的关联方法研究[D]. 北京:中国科学技术信息研究所,2016.
[24] W3SCHOOL. RDF[EB/OL]. (2012-09-28) [2023-02-18]. http://www.w3school.com.cn/rdf/rdf_intro.asp.
[25] BERNERS-LEE T,FIELDING R,MASINTERL. Uniform resource identifiers (URI):generic syntax[EB/OL]. (1998-08-01) [2023-02-18]. http://www. ietf.org/rfc/rfc2396.txt.
[26] 国务院学位委员会,教育部.研究生教育学科专业目录（2022 年）[EB/OL]. (2022-09-13) [2023-03-15]. http://www.gov.cn/zhengce/zhengceku/2022-09/14/5709785/files/bb23485679bf489aba5d3299c36e0606.pdf.
[27] 国家图书馆《中国图书馆分类法》编辑委员会.中国图书馆分类法(第五版)[M].北京: 国家图书馆出版社,2010.
[28] Clarivate. Web of Science Research Areas [EB/OL]. (2021-12-01) [2023-03-15]. https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm?Highlight=research%20area.

作者贡献说明：

梁美宏：论文撰写与修改；
胡晓辉：论文指导与建议；
束彭飞：理论与实践现状调研分析；
刘飞：实证数据清洗与分析。

Research on Institutional Repository Entity Academic Relationship Discovery Method Based on Linked Data

Liang Meihong[1] Hu Xiaohui[1] Shu Pengfei[1] Liu Fei[2]

(1. Library of Hohai University Nanjing Jiangsu 210098

2. Library of Nanjing Tech University Pujiang Institute Nanjing Jiangsu 211200)

Abstract：[Purpose/significance] By using the uniqueness and linkability of linked data, the paper tries to discover and correlate academic relationship of institutional repository entities，in order to realize institutional repository resources semantic aggregating. **[Method/process]** It analyzes the types and characteristics of the entity academic relationship and linked data requirements, then puts forward academic relationship association methods. Finally it uses the data of Hohai University Institutional Repository to conduct an empirical study.

[Results/Conclusion] The paper builds the method of discovering entity academic relationships based on linked data. It finds out the academic relationships of empirical data and points out its application value. This method can realize semantic associating of institutional repository resources, and provide knowledge retrieval service based on linked data, which can meet user's knowledge needs.
Keywords：institutional repository; linked data; academic relation; discovery method